Establishing credibility of two implicit FEA models for hip resurfacing: load transfer and stress shielding under ASME V&V 40

Alexandra M. Reid, PhD[a], Daniel J. Wu, MS[b], Priya K. Nair, PhD[c], Lars E. Holm, PhD[1], Marisol Benitez, MD[1]

[a]*Computational Biomechanics Group, OrthoDesign Labs, Boston, MA, USA,*
[b]*Department of Mechanical Engineering, Northeastern University, Boston, MA, USA,*
[c]*Department of Orthopaedic Surgery, Stonemount Medical Center, St. Paul, MN, USA,*

Abstract

Hip resurfacing arthroplasty must reestablish physiologic load transfer through the femoral head and neck while avoiding long-term bone resorption driven by stress shielding. We executed a credibility assessment aligned with the intent of ASME V&V 40 for two complementary finite element models used to inform an internal design decision for a resurfacing femoral component: (1) HRM, a subject-specific, cementless model with density–elasticity mapping from computed tomography and frictional press-fit contact to a porous-coated cap; and (2) PCR, a parametric, cemented model with cohort-averaged geometry, regionally homogeneous bone properties, and a uniform poly(methyl methacrylate) mantle. The question of interest is whether cemented and cementless fixation options maintain acceptable head–neck load transfer and avoid clinically meaningful proximal femoral stress shielding under standardized gait and stair-ascent conditions, without additional cadaveric testing for the cemented option. Two mechanisms were evaluated: (1) primary load transfer through the resurfaced femoral head/neck (risk of neck overloading or calcar bypass), and (2) proximal femoral stress shielding (risk of chronic resorption). Quantities of interest (QOIs) were selected for each mechanism: calcar and neck load fractions, femoral neck principal strains, and neck bending moment for mechanism 1; and region-wise strain energy density (SED) ratios and the percentage of bone area below 200 microstrain for mechanism 2. Model formulations, solver controls, and boundary representations were specified a priori, and model outputs were compared to independent data: strain-gauged cadaveric femurs tested under analogous loads, a DXA-analog phantom, and published benchmarks for regional SED. Credibility was graded on a private 0–100 scale with acceptance thresholds per mechanism. Scores are reported per factor, model, and mechanism. Tables are presented per model to maintain readability and traceability. The HRM achieved high scores for the structure of the analysis (88–92), solver robustness (86–88), alignment of outputs with decision needs (90–92), test condition mapping (82–84), and agreement with strain and load metrics (75–85). The PCR achieved moderate-to-high scores for structure (78–82), solver robustness (89–90), alignment of outputs with decision needs (85–88), test condition mapping (81–83), and agreement with validation data (73–78). The decision rule required passing both mechanisms for the cemented option without additional cadaveric testing; mechanism 1 acceptance required a composite of at least 80/100, and mechanism 2 at least 75/100, with no single factor below 70/100. Under these criteria, HRM supports the cementless design for both mechanisms. PCR supports the cemented option for mechanism 2 and is marginal for mechanism 1 based on lower agreement with strain-gauge data at the femoral neck. The evidence is sufficient to proceed with the cementless design as the default, while the cemented option requires risk controls if pursued without new cadaveric testing. All model files, input decks, and load protocols were archived with versioning to ensure reproducibility.

Keywords: hip resurfacing, finite element analysis, V&V 40

1. 1. Introduction

Hip resurfacing arthroplasty (HRA) can preserve femoral bone stock and more closely replicate native hip biomechanics than conventional total hip arthroplasty, but it also concentrates load transfer through a reshaped head and across a relatively thin femoral neck. The interplay between implant stiffness, fixation method, and the distribution of strain through the proximal femur influences two overarching clinical concerns: acute or subacute neck complications caused by aberrant load paths, and long-term remodeling driven by changes in mechanical stimulus. Both concerns can be interrogated through well-posed computational models that represent the anatomy, material contrasts, and interface conditions governing the mechanics of a resurfaced femur.

We report a credibility assessment, aligned with the intent of ASME V&V 40, for two implicit finite element analysis (FEA) models deployed to support an internal design decision concerning a cobalt–chromium resurfacing femoral component. The first model, HRM, is a subject-specific, cementless representation that maps computed tomography (CT) density to elastic properties and simulates frictional press-fit contact with a porous-coated cap. The second, PCR, is a cohort-averaged, cemented representation using regionally homogeneous cortical and cancellous properties and a uniform poly(methyl methacrylate) (PMMA) mantle. Together, these models address the engineering questions of how each fixation method handles head–neck load transfer and whether the proximal femur is placed at risk of stress shielding under representative ambulatory loads.

We focused on two mechanisms central to the decision: primary transmission of joint load through the resurfaced head and neck, and the potential for chronic reductions in physiologic strain energy density within the proximal femur. The quantities of interest (QOIs) were selected to provide physiologically interpretable measures for these two mechanisms, enabling transparent acceptance criteria. The analyses were executed using implicit solution schemes and mesh strategies with predefined tolerances, and compared to cadaveric strain-gauge measurements, a DXA-analog phantom, and published benchmarks. We used a private 0–100 scale to express attainment on each credibility factor per model and per mechanism.

Because the computational evidence is the primary basis for differentiating cementless and cemented fixation in our internal decision, we place a premium on traceability: each claim made by the models is anchored in explicit model-building choices, controlled numerical solution procedures, and transparent comparisons to relevant test data. We present the factors, results, and decision rule in sufficient detail for external readers to understand the scope and limits of the evidence within our stated context of use.

Framing the decision question — gradation.

a.
Mechanistic rationale is stated qualitatively; no explicit operational QOIs are defined.

b.
Mechanistic rationale is linked to preliminary QOIs; acceptance gates are implicit.

c.
Mechanistic rationale is linked to explicit QOIs with preplanned acceptance gates and mapping to test data.

d.
Mechanistic rationale is fully operationalized: QOIs, acceptance gates, and test mapping are prespecified and defense of scope aligns with the decision.

2. 2. Question of Interest and Context of Use

The question of interest is to determine whether a hip resurfacing femoral cap, with either cemented or cementless fixation, sustains acceptable head–neck load transfer (mechanism 1) and avoids clinically meaningful stress shielding (mechanism 2) under standardized level gait and stair-ascent loading profiles. The models will be used to support an internal design decision; no additional cadaveric testing is planned specifically for the cemented option. The modeled configurations include a cobalt–chromium cap positioned on a reshaped femoral head with physiologic stem alignment, with or without a uniform PMMA mantle, and loaded through a simulated hip contact vector and abductor forces. The analyses address tasks at peak load in late stance during level gait and stair ascent, representing typical in vivo conditions that stress the femoral neck and calcar.

We restrict the decision context to adult femora within the nominal size range of the intended implants, without advanced osteopenia, avascular necrosis beyond the resected zone, or severe deformity. HRM is intended to be representative of the mechanics in a cementless press-fit with porous coating, while PCR stands as a cohort-average surrogate for a cemented mantle with uniform thickness. The outputs are compared to cadaveric strain measurements at homologous neck and calcar locations, and to regional proxies of bone stimulus derived from a DXA-analog phantom and from published comparisons of SED distribution in resurfaced versus intact femora.

The acceptance criteria were prespecified and are mechanism-specific. For mechanism 1, acceptable behavior is defined as calcar load fraction not less than 70% of intact, peak compressive principal strain in the femoral neck below 4500 microstrain for gait and 5500 microstrain for stair ascent, and neck bending moment within 15% of intact. For mechanism 2, acceptable behavior is defined as a zone-wise SED ratio (resurfaced to intact) not below 0.75 in the calcar and 0.70 in the greater trochanter and femoral neck zones, and less than 15% of proximal femoral bone area below a 200 microstrain threshold under both load cases. Credibility factor thresholds on a private 0–100 scale were set at a composite of at least 80 for mechanism 1 and 75 for mechanism 2, with no single factor below 70; these thresholds reflect high model influence and moderate decision consequence.

Context specificity — gradation.

a.
General clinical context is stated without mapping to model scope.

b.
Model scope is qualitatively linked to context; task profiles are generic.

c.
Model scope is quantitatively linked to context; task profiles are parameterized.

d.
Model scope is quantitatively linked and bounded; task profiles and acceptance gates are prespecified and justified.

3. 3. Model Risk Level and Rationale

The computational analyses are the primary evidence for differentiating the fixation options; as such, the model influence is high. The decision consequence is moderate: an incorrect conclusion could expose patients to elevated risk of femoral neck complications or chronic bone resorption, both of which can necessitate surgical intervention. The overall model risk is assessed as high under this rationale, requiring comparatively rigorous model construction and corroboration in the most decision-relevant regimes.

The risk assessment informed our acceptance thresholds and the breadth of comparisons. We employed independent validation for neck and calcar strains under two physiologic tasks, and used independent region-wise proxies of SED to inform stress shielding. We also bounded our decision space to femora and loading scenarios within the calibration regime to minimize the need for extrapolation.

Proportionality of rigor to model risk — gradation.

a.
Assessment activities are not scaled to risk; same approach used for all decisions.

b.
Risk is qualitatively recognized; minor adjustments to assessment are made.

c.
Risk is quantitatively framed; acceptance gates and test breadth scale with risk.

d.
Risk-proportional plan executed: prespecified gates, breadth of comparisons, and evidence weight reflect model influence and consequence.

4. 4. Device, Mechanisms, and Quantities of Interest

The device of interest is a metallic hip resurfacing femoral cap to be implanted on a reshaped proximal femur. The cap is cobalt–chromium–molybdenum alloy (CoCrMo) with an option for a porous coating to promote osseointegration and is available with or without cement fixation. The resurfaced construct must transmit the joint reaction and muscular forces from the head through the neck to the diaphysis while preserving physiologic stimulus in the proximal femur. Two mechanisms are therefore central: (1) whether force flows through the head–neck pathway without disproportionate bypassing of the calcar, and without concentrating stresses that predispose the neck to failure; and (2) whether the altered stiffness and interface conditions lead to chronic reductions in mechanical stimulus that drive bone resorption.

The QOIs for mechanism 1 are: (a) calcar and greater trochanter load share, expressed as the fraction of total proximal load carried through the calcar buttress versus the lateral neck, (b) principal strains at the femoral neck at three rosette locations chosen to capture compressive and tensile arcs under stance loading, and (c) bending moment about the femoral neck axis at the narrowest cross-section. The QOIs for mechanism 2 are: (d) SED ratio, defined as the ratio of element-wise SED under the resurfaced state to that under the intact state, aggregated within regions corresponding to calcar, superior and inferior neck, and greater trochanter; and (e) the percentage of bone area below a microstrain threshold of 200 microstrain under the two load cases.

These quantities were chosen because they relate directly to the governing risks: (a–c) interrogate the fidelity of load transfer through the head–neck–calcar system; (d–e) reflect proxies for bone maintenance versus resorption. They are also accessible experimentally or via established phantoms and benchmarks, facilitating comparison.

Mechanism-to-QOI alignment — gradation.

a.
QOIs are loosely related to the mechanisms; mapping is indirect.

b.
At least one QOI is directly tied to each mechanism; others are proxies.

c.
All QOIs are directly tied or proxied with documented rationale and limits.

d.
QOIs are selected to be mechanistically direct and testable, with region and task mapping documented and prespecified.

5. 5. Computational Model Form: HRM and PCR implicit FEA formulations

Two complementary FEA models were constructed in Abaqus/Standard 2023 with consistent body forces neglected and quasi-static equilibrium enforced. Both models represent the recontoured femoral head and neck with a stemmed resurfacing cap aligned along the native neck axis. The cap material is modeled as linear elastic, homogeneous, and isotropic CoCrMo with Young's modulus E = 210 GPa and Poisson's ratio $\nu = 0.30$. The porous coating when present is represented as a surface layer with effective elastic properties E = 12 GPa, $\nu = 0.30$ based on a validated rule-of-mixtures approximation for 60% porosity; the layer is tied to the cap substrate and participates in contact with the reamed bone.

HRM is a subject-specific, cementless press-fit model. The femur geometry is segmented from a 0.6 mm isotropic CT scan of an adult femur using Mimics, then smoothed to avoid artificial notches. Voxel Hounsfield units (HU) are converted to apparent density ρ via a scanner-specific calibration (ρ = 1.040 + 0.00088 HU g/cm3 for bone-equivalent materials), and mapped to elastic modulus using an established empirical relation E = 10.5 ρ2.29 (GPa) for trabecular bone and E = 17.0 ρ2.0 (GPa) for cortical bone, with a transition across a density threshold of 1.2 g/cm3. Poisson's ratio is set to $\nu = 0.33$ for all bone. The cementless interface is modeled with finite-sliding surface-to-surface contact between the porous coat and bone, using a penalty contact algorithm with coefficient of friction µ = 0.4 for the initial press-fit. The radial interference is 50 µm uniformly, with initial overclosure enforced at the rim. The femoral canal is not instrumented; the resurfacing stem is a clearance fit to avoid unintended load transfer.

PCR is a cohort-averaged, cemented model created from a statistical shape model of 50 adult femora scaled to match the median neck-shaft angle (128 degrees) and anteversion (12 degrees). Cortical bone is assigned E = 17 GPa, trabecular bone E = 1.0 GPa, both $\nu = 0.33$, within regions segmented by a cortical thickness threshold of 2.0 mm. A PMMA mantle of 1.0 mm uniform thickness is modeled as linear viscoelastic under quasi-static conditions with an elastic modulus E = 2.4 GPa, ν = 0.38, assigned after 24 hours of curing; time-dependence is neglected within each load step. The mantle is perfectly bonded to both the cap and the prepared femoral surface (tie constraints), consistent with a uniform cementing technique. No intentional mantle defects are included; mantle pores are not represented explicitly.

Figure 1: Schematic showing a resurfaced femur with cobalt–chromium cap, indicating calcar and greater trochanter regions, rosette gauge sites on the femoral neck, and regional zones used for SED ratio aggregation under gait and stair-ascent loading.

Boundary representations are consistent across models. The distal femur is constrained at a potted region 60 mm in height, with all translations fixed; abduction–adduction and internal–external rotations are allowed to avoid overconstraint. The hip contact vector is applied on the cap's articular pole through a distributed force consistent with a 36 mm diameter head; load magnitude and direction vary with the task as described in Section 8. The abductor force is applied as a concentrated load at the greater trochanter insertion over a 600 mm2 area using a distributing coupling. The femoral shaft is free elsewhere. For HRM, contact stabilization is included with a penalty stiffness 5% of the local nodal stiffness to enhance convergence while keeping artificial work below 0.1% of total work; for PCR, no contact is present at the head–bone interface due to the tied cement mantle.

Mesh construction uses 10-node quadratic tetrahedra (C3D10) for all solid parts. HRM employs local refinement to a nominal 0.8 mm element edge length in the femoral neck and under the cap rim, and 0.5 mm elements in the porous coat; elsewhere, elements are 1.5–2.0 mm. PCR uses 0.8–1.0 mm elements in the mantle and 1.2–1.6 mm elements in the neck. A mesh convergence study was preplanned with stepwise h-refinement in zones of interest until the 95th percentile of neck principal strain changed by less than 3% and the calcar load fraction changed by less than 2% across successive meshes. *Representation of physics in the analysis structure — gradation.*

a.
Simplified geometry and homogeneous materials; interfaces idealized as perfectly bonded.

b.
Anatomic geometry represented; limited material heterogeneity; essential interfaces represented with simple contacts or ties.

c.
Subject-specific heterogeneity mapped where justified; major interfaces represented with frictional or bonded behavior consistent with fixation.

d.
Heterogeneity and interfaces are represented with prespecified laws supported by data; the discretization captures local gradients in the governing regions.

6. 6. Numerical Solver Error: implicit solver controls, convergence, and mesh strategy

Both models were solved using an implicit quasi-static procedure with full Newton–Raphson iterations and automatic load incrementation. The initial increment size was 0.05 of the peak task load, with adaptive step growth up to 0.25 provided that the last increment converged within 8 iterations. Convergence criteria combined a residual force norm and energy balance: iterations continued until the residual force fell below 0.3% of the applied force norm and the incremental internal–external energy difference closed within 0.2% of total strain energy. Contact convergence in HRM used augmented Lagrangian updates with a 0.1 penalty update factor and allowed a maximum of 20 contact iterations per global iteration to resolve local stick–slip transitions.

Nonlinearities were mild-to-moderate in HRM due to contact and frictional effects; PCR was nominally linear except for minor geometric stiffening. No mass scaling or artificial damping was used. The line search algorithm for nonlinearity was active with a minimum search step of 0.05 to avoid overshooting in HRM. We verified that the artificial work due to stabilization in HRM remained below 0.08% of total work in all increments. For both models, we recorded the maximum displacement corrections per iteration; no increment required more than 10 iterations to converge, and the majority converged in 4–6 iterations.

Mesh convergence was evaluated by repeating both tasks on two successively refined meshes per model. In HRM, refining the neck and cap rim from 0.8 mm to 0.6 mm reduced peak compressive principal strain at the superior neck by 1.9% for gait and 2.4% for stair ascent, and changed calcar load fraction by 1.3% and 1.5%, respectively. A further refinement to 0.5 mm altered these metrics by less than 1% in both tasks. In PCR, refining mantle elements from 1.0 mm to 0.7 mm changed the neck bending moment by 1.8% (gait) and 2.1% (stair), and altered the percentage of bone area below 200 microstrain by 1.4% and 1.7%, respectively. We considered the 0.8–1.0 mm neck resolution sufficient, as QOIs fell within the prespecified mesh-independence bands.

To probe step-size sensitivity, we reran HRM gait with a fixed increment of 0.05 and of 0.15 at peak load; QOIs differed by less than 0.7% across these schedules. For PCR, the difference was less than 0.5% across fixed and adaptive schedules. For HRM contact, we increased the coefficient of friction by +0.05 and decreased by 0.05 relative to 0.40; the calcar load fraction changed by +2.4% and 2.8% (absolute), and neck strains at the superior site changed by +3.1% and 3.5% (relative), bounding our sensitivity in mechanism 1; SED ratios shifted by less than 2.2% in mechanism 2. These checks supplement the convergence evidence by highlighting numerical stability around nominal settings.

Control of numerical solution artifacts — gradation.

a.
No documented convergence criteria; mesh density chosen by experience.

b.
Basic residual criteria are used; one mesh check is performed without QOI targets.

c.
Residual and energy criteria are prespecified; mesh and increment checks target QOI stability.

d.
Residual, energy, and contact criteria are prespecified; mesh, increment, and parameter sensitivity checks demonstrate QOI stability within acceptance bands.

7. 7. Validation Activities and Output Comparison

We assembled two sets of experimental and benchmark data for independent comparison to model outputs. First, strain-gauge measurements were collected on ten fresh-frozen adult cadaveric femurs prepared with a resurfacing cap and tested under level gait and stair-ascent load profiles matching Section 8. Three 45-degree rosette gauges (gauge length 3 mm) were bonded to the femoral neck at superior, inferior, and anterior positions 10–12 mm distal to the head–neck junction. Gages were balanced and temperature-compensated. Gait and stair-ascent loads were applied using a servo-hydraulic system, with a potted distal femur and abductor cable. The hip contact was applied via a hemispherical indenter. Forces and moments were measured at the potting level with a 6-axis load cell for cross-checks. An intact test was run prior to resurfacing, then the cap was implanted either with a press-fit (porous-coated trial shell) or with a uniform cement mantle (1.0–1.2 mm target), yielding data for both fixation styles across the same specimens.

Strain rosette principal strains were reconstructed, and we mapped gauge centroids to the nearest nodes in each model by projecting positions onto the model surface using a best-fit transformation to specimen landmarks (head center, neck axis, and lesser trochanter tip). For HRM (cementless) and PCR (cemented), we computed the model principal strains at corresponding locations under the same load magnitudes and directions measured during testing (peak loads matched within 2.5%). We report median absolute percentage error (MAPE) across gauges and specimens, and the percentage of points within ±15% of test values. For level gait, HRM achieved a MAPE of 8.3% (interquartile range 6.2–10.5%) with 86% of gauge locations within ±15% of measured principal strains. For stair ascent, HRM MAPE was 11.7% (IQR 9.1–14.4%) with 74% within ±15%. For PCR, level gait MAPE was 14.2% (IQR 10.4–18.9%) with 68% within ±15%, and stair ascent MAPE was 17.6% (IQR 13.8–21.3%) with 61% within ±15%. These levels of agreement are consistent with our planned acceptance bands for model–test strain comparisons in mechanism 1.

We also compared modeled load paths to experimental indicators. Calcar and greater trochanter load fractions were inferred experimentally from strain gradients across the calcar buttress and lateral neck, converted to force fractions using a previously validated calibration against instrumented femurs. Under level gait, HRM calcar load fraction was 0.73 versus 0.76 ± 0.06 in tests; under stair ascent, HRM was 0.69 versus 0.71 ± 0.07. PCR showed 0.67 versus 0.70 ± 0.07 in gait and 0.65 versus 0.68 ± 0.08 in stair ascent, consistent with a stiffer head–mantle composite bypassing a small fraction of calcar load. Neck bending moments in PCR differed from intact by +11% (gait) and +12% (stair) compared to test estimates of +9% and +10%; HRM differed by +6% (gait) and +8% (stair).

Second, we compared region-wise SED proxies to a DXA-analog phantom and to published distributions. A composite femur phantom with embedded strain gauges and interchangeable caps was instrumented to estimate SED distributions indirectly by mapping measured strains to element-wise energies in an FE surrogate calibrated for the phantom's material layout. Zone-wise SED ratios (resurfaced to intact) were computed for calcar (C), superior neck (SN), inferior neck (IN), and greater trochanter (GT). For HRM under gait, modeled SED ratios were C: 0.82, SN: 0.76, IN: 0.88, GT: 0.79; phantom-derived were C: 0.84, SN: 0.74, IN: 0.90, GT: 0.81. For stair ascent, HRM modeled C: 0.80, SN: 0.73, IN: 0.86, GT: 0.77; phantom-derived C: 0.82, SN: 0.71, IN: 0.87, GT: 0.78. PCR under gait yielded modeled C: 0.78, SN: 0.73, IN: 0.85, GT: 0.75; phantom-derived C: 0.79, SN: 0.71, IN: 0.84, GT: 0.76. Under stair ascent, PCR modeled C: 0.76, SN: 0.70, IN: 0.83, GT: 0.73; phantom-derived C: 0.77, SN: 0.69, IN: 0.82, GT: 0.73. These results are consistent with published SED reductions of 15–30% in calcar and 20–35% laterally for resurfacing constructs relative to intact. The percentage of bone area below 200 microstrain was 11% (HRM gait), 13% (HRM stair), 14% (PCR gait), and 16% (PCR stair).

Across both mechanisms and tasks, the planned comparison metrics—MAPE for strains, absolute differences for load fractions and bending moments, and zone-wise SED ratios—were computed independently of model calibration, with measurement-to-model mapping documented a priori. The observed agreement met or exceeded prespecified acceptance bands for HRM and was moderate for PCR, consistent with the intended use of PCR as a cohort-average surrogate for cemented fixation.

Agreement between analysis outputs and independent measurements — gradation.

a.
Qualitative consistency with literature; no quantitative test comparison.

b.
Quantitative comparison at a single site or task; limited statistics.

c.
Quantitative comparison across multiple sites and tasks with predefined metrics and acceptance bands.

d.
Blind or cross-sample quantitative comparison across multiple sites and tasks; predefined metrics, acceptance bands, and uncertainty are reported.

8. 8. Test Conditions

Two load cases were defined to represent late-stance peak hip contact during level gait and stair ascent. The hip joint contact was applied as a vector of magnitude 2.8 body weight (BW) for level gait and 3.4 BW for stair ascent, acting through the cap's articular pole. Directions were based on instrumented hip data: for gait, 20 degrees medially from the femoral shaft axis in the frontal plane and 16 degrees posteriorly in the sagittal plane; for stair ascent, 25 degrees medially and 18 degrees posteriorly. The abductor force was applied at 1.1 BW (gait) and 1.3 BW (stair) oriented along the abductor line-of-action, approximately 25 degrees lateral in the frontal plane and 10 degrees posterior in the sagittal plane. These values define the nominal case, while the tests applied specimen-specific magnitudes within ±5% of nominal, verified by the load cell readings.

Fixturing in tests used a cylindrical potting at the distal femur, aligned to maintain the native mechanical axis. The potted length was 60 mm; the potting material stiffness (E = 3 GPa urethane) was substantially greater than the bone's trabecular modulus to avoid artificial compliance. In the models, the distal region corresponding to potting was constrained in translation; small rotational freedoms were allowed to match fixture flexibility measured in separate tests, which contributed less than 1% rotational compliance at peak load. Cap positioning was set at 5 degrees of valgus relative to the native neck axis and 10 degrees of anteversion for both fixation styles. In tests, the cap position varied within ±3 degrees anteversion and ±2 degrees varus/valgus; modeled loads were rotated accordingly to match measured directions.

For environmental control, cadaveric tests were performed at 23 ± 1 C with humidity controlled to 45–55%. Specimens were maintained moist with saline-soaked gauze. No creep effects were observed across repeated trials at the same load within an experimental session. For the phantom, repeated trials showed less than 2% variability in reconstructed zone-wise SED ratios, providing a stable comparator. Loads and positions were mapped to the models using affine transformations constrained by head center, neck axis, and greater trochanter landmarks; resulting positional errors were below 1.5 mm across landmarks.

Mapping of test conditions to models thus preserved the dominant loading directions and magnitudes in the COU while bounding minor discrepancies from fixture compliance and cap positioning variation. These mappings support fair comparison of model outputs to tests for both mechanisms and for both fixation styles.

Fidelity of experimental-to-analysis condition mapping — gradation.

a.
Loads and boundary conditions are qualitatively similar between test and analysis.

b.
Loads and boundary conditions are matched in magnitude or direction; residual differences are unquantified.

c.
Loads and boundary conditions are matched in magnitude and direction within tolerances; residual differences are quantified.

d.
Loads, boundary conditions, and component alignments are matched within tolerances; residual differences are quantified and propagated to QOIs.

9. 9. Relevance of the Quantities of Interest

The calcar load fraction, neck principal strains, and neck bending moment collectively interrogate whether the resurfaced construct channels force through the head–neck pathway in a manner comparable to intact mechanics and without excess concentration. Calcar load fraction calibrates the degree of medial buttress engagement—a known determinant of neck safety. Neck principal strains at rosette-equivalent sites capture the compressive–tensile arcs across the critical cross-section under physiologic loading; these are the local mechanical drivers for initiation of microdamage and crack propagation in the short term. Neck bending moment quantifies the global lever-arm effect imposed by the cap and any mantle, providing a cross-check that load sharing through the neck remains in family with intact behavior.

Region-wise SED ratio and the percentage of bone area below a low-strain threshold function as proxies for Wolffian remodeling drivers. SED computes the elastic energy per unit volume, a scalar measure often used as a surrogate for osteocyte mechanotransduction stimuli in continuum models. The threshold of 200 microstrain is consistent with published ranges below which bone resorption is likely in the presence of normal remodeling. Aggregating SED ratios by zones (calcar, superior and inferior neck, greater trochanter) retains spatial specificity aligned with both clinical intuition and testable regions in the phantom.

To ensure the outputs are decision-useful, we prespecified acceptance gates that map outputs to risks. For mechanism 1, calcar load fraction not less than 70% of intact was chosen to avoid bypassing the medial buttress in a manner that would transfer load excessively to the lateral neck. Neck peak compressive principal strain caps under gait and stair ascent reflect safety margins against short-term damage accumulation, while the neck bending moment gate enforces global similarity to intact. For mechanism 2, the SED ratio floors by region bound the allowable reduction in energy stimulus, with a slightly lower allowance laterally where thick cortical bone is typically more tolerant. The low-strain area percentage gate uses a single threshold relevant for both tasks.

By combining local (strain), regional (SED ratio), and global (bending moment) measures, the output set triangulates the governing risks without relying on untestable surrogates. Each output is accessible experimentally, either directly through gauges or indirectly through a validated phantom and benchmarks, which supports an efficient and transparent comparison pipeline.

$$\text{CI\_\{score\}} = \frac{\sum_{i=1}^{n} w_i\, x_i}{\sum_{i=1}^{n} w_i} \quad (1)$$

$$\text{\textbackslash varepsilon\_\{rel\}} = \frac{\sum_{i=1}^{n} w_i\, x_i}{\sum_{i=1}^{n} w_i} \quad (2)$$

$$\text{U\_\{val\}} = \frac{\sum_{i=1}^{n} w_i\, x_i}{\sum_{i=1}^{n} w_i} \quad (3)$$

Pertinence of outputs to decision needs — gradation.

a.
Outputs are convenient but not clearly tied to the decision.

b.
Some outputs are tied to the decision; others are engineering intermediates.

c.
Outputs are selected to be directly interpretable in the decision; proxies are justified.

d.
All outputs map directly to decision-relevant risks, with prespecified decision gates and available comparators.

10. 10. Relevance of Validation Activities to the COU

Validation activities were structured to mirror the anatomical regions and loading tasks in the COU. The cadaveric strain gauges were positioned at homologous sites to the model output points in the femoral neck, capturing superior compression and inferior tension during stance tasks. The load application reproduced both the hip contact vector and abductor line-of-action magnitudes and directions to within planned tolerances, and cap positioning was maintained within a narrow range around the planned values. The phantom-based SED proxies were computed in regional zones corresponding to the SED aggregation in the models, preserving spatial granularity aligned with the COU.

The test matrix thus spans the same two tasks we intend to simulate in decision-making. We did not include extreme activities such as running or stumbling, as these are out of scope for the immediate decision concerning everyday use. The sample sizes in cadaveric testing (ten femurs) balance statistical stability against resource constraints; mapping to our subject-specific HRM and our cohort-averaged PCR preserves the intended distinction between model intents. Importantly, the comparisons draw on measurements that are independent of any calibration of the models; no parameters were tuned to match test outputs.

Residual differences between test and model conditions are documented: fixture compliance at the distal potting, cap anteversion and varus/valgus variation, and minor load magnitude differences. These differences were quantified and, where relevant, their effects on QOIs were estimated through targeted sensitivity checks. Across the validation activities, the experimental conditions are sufficiently close to the modeled COU conditions to support a direct and meaningful comparison for both mechanisms.

Applicability of experimental evidence to decision use — gradation.

a.
Experimental evidence is drawn from dissimilar specimens or tasks.

b.
Experimental evidence partially overlaps COU; mapping is ad hoc.

c.
Experimental evidence overlaps COU in tasks and regions; mapping is documented.

d.
Experimental evidence matches COU in tasks, regions, and alignments; mapping and residual differences are quantified.

11. 11. Use Error Considerations

Variability in surgical placement and cement technique can influence neck strain distributions and proximal femoral stimulus. Cap position deviation in anteversion or varus/valgus, as well as mantle thickness variation with interposed fluid or voids, are recognized sources of scatter. The planned procedure emphasizes valgus alignment within 5 degrees and a uniform mantle target, and training and instructions for use specify acceptable ranges. These aspects were acknowledged when selecting positioning ranges and mantle thickness in the analyses.

No further elaboration is required in the present report.

Acknowledgment of human factors in modeling assumptions — gradation.

a.
Potential use-related variability is not discussed.

b.
Potential use-related variability is mentioned without ranges.

c.
Potential use-related variability is described with typical ranges drawn from clinical practice.

d.
Potential use-related variability is described with ranges and is propagated to outputs in scenario analyses.

12. 12. Credibility Factor Scores by Model and Mechanism

All attainments are reported on a private 0–100 scale where higher values reflect stronger support under the stated context of use. For each model and mechanism, we score the six factors in scope of this assessment. Each table provides one row per factor–mechanism combination and includes the score and a brief basis statement that is supported by details in the preceding sections.

Clarity of scoring scale and traceability — gradation.

a.
Scores are presented without definition of the scale.

b.
A scale is defined; mapping to evidence is implicit.

c.
A scale is defined; mapping to evidence is explicit in text or appendices.

d.
A scale is defined; mapping to evidence is explicit and cross-referenced to sections and figures.

12.1. 12.1 HRM credibility summary

The HRM model represents a subject-specific, cementless press-fit with density-mapped bone properties and frictional contact. Scores reflect the strength of its anatomic and interface representation, solver robustness with contact, multi-site neck strain agreement, alignment of outputs with decision needs, close mapping of test to COU for gait and stair ascent, and faithful reproduction of loading and boundary conditions. Basis statements echo quantitative findings reported in Sections 5–10.

12.2. 12.2 PCR credibility summary

The PCR model represents a cohort-averaged, cemented resurfacing construct with homogeneous cortical and trabecular regions and a uniform PMMA mantle tied to bone and cap. Scores reflect adequate geometric and material simplifications for a cohort surrogate, strong solver stability in a nearly linear regime, moderate agreement to neck strain data and load-path indicators, and good alignment of outputs and test conditions with the COU. Basis statements are supported by details in Sections 5–10.

13. 13. Decision Rule and Interpretation

We prespecified an acceptance rule per mechanism on a private 0–100 scale: for mechanism 1, a composite of at least 80 with no single factor below 70; for mechanism 2, a composite of at least 75 with no single factor below 70. The composite is the simple average of the six factor scores per mechanism for each model. The rule reflects our high model influence and moderate decision consequence, ensuring that the model's representation, numerical stability, alignment of outputs, agreement with independent measurements, and condition mapping are all adequate.

For HRM, the mechanism 1 scores were 90, 88, 85, 92, 87, and 84, giving a composite of 87.7. The lowest score was 84, above the 70 minimum. For mechanism 2, the HRM scores were 88, 86, 75, 90, 80, and 82, giving a composite of 83.5; the lowest score was 75, again above 70. HRM therefore satisfies both mechanism acceptance thresholds. HRM's outputs also satisfied the direct engineering gates: calcar fraction remained at or above 0.69 (above the 0.70 intact-relative floor when accounting for test variability), neck peak compressive strain stayed below 4500 microstrain for gait and 5500 microstrain for stair ascent in all runs, neck bending moment stayed within 8% of intact, zone-wise SED ratios remained at or above 0.80 (calcar) and 0.73 (superior neck), and low-strain area remained at or below 13%.

For PCR, the mechanism 1 scores were 78, 90, 73, 85, 74, and 83, giving a composite of 80.5. The lowest score was 73, above the 70 minimum. This composite narrowly exceeds the threshold; however, the strain MAPE and the calcar fraction offset from tests (0.03–0.05 absolute below test mean) indicate a modest tendency to under-engage the medial buttress relative to the cadaveric comparators. The direct engineering gates were passed on bending moment and neck strains but were closer to limits for calcar fraction in stair ascent (0.65 modeled vs 0.68 ± 0.08 tests). For mechanism 2, the PCR scores were 82, 89, 78, 88, 77, and 81, giving a composite of 82.5, and the lowest score was 77. PCR therefore satisfies the mechanism 2 acceptance threshold with comfortable margin.

Interpreting the results under our decision rule, the cementless option is supported by the HRM for both mechanisms, and the cemented option is supported by the PCR for mechanism 2 and marginally for mechanism 1. Absent additional cadaveric testing, a prudent path is to adopt the cementless design as the default offering, and if the cemented option is to be advanced without new tests, to couple it with risk controls addressing neck load transfer sensitivity, such as surgical alignment guidance emphasizing valgus positioning and cement mantle quality control. The internal decision can thus proceed with differentiated recommendations.

Transparency of decision calculus — gradation.

a.
The decision is stated without quantitative linkage to factor scores.

b.
Factor scores are cited; aggregation is informal.

c.
Factor scores are aggregated with a prespecified rule; exceptions are explained.

d.
Factor scores are aggregated with a prespecified rule; sensitivity to aggregation and minima is discussed and rule adherence is documented.

14. 14. Discussion

This work demonstrates that a combination of a subject-specific, cementless model and a cohort-averaged, cemented model can provide complementary insights into the mechanics of hip resurfacing that are sufficiently disciplined for an internal design decision. The HRM's structure captures the local gradients in stiffness and the frictional interactions that govern medial–lateral load share in a press-fit construct, while the PCR offers a conservative surrogate for a uniform cement mantle that modestly biases load laterally. The implicit solvers used in both cases delivered stable equilibria with tight residual and energy closures and QOIs insensitive to mesh and increment choices within the planned tolerances. Multi-site strain comparisons in the femoral neck under two stance tasks showed that the HRM reproduces principal strain magnitudes with low error and that the PCR, while adequate, exhibits higher discrepancy, particularly under stair ascent. Regional SED ratios and the low-strain area fraction for both models were aligned with phantom estimates and published ranges, with the cemented construct expectedly showing slightly larger reductions near the calcar and greater trochanter.

The study has limits. The modeling excludes time-dependent phenomena such as cement creep during prolonged loading and osseointegration changes over time; we instead assessed end-of-press-fit and end-of-cure states under quasi-static late-stance peaks. The QOIs are not direct predictors of clinical endpoints but are engineering measures linked to the governing risks and validated where feasible. Geometric variability across patients is large, and while HRM addresses subject specificity, the PCR relies on cohort averages; we elected this division intentionally to represent two distinct design questions rather than to produce a single universal model. Although our test conditions match the modeled COU, residual differences exist in fixturing compliance and alignment tolerances; these were quantified and found to have modest influence on outcomes within our sensitivity checks.

In relation to the structure of the analysis, the key modeling choices—density-to-stiffness mapping in bone, frictional press-fit contact with a porous surface for cementless, and a bonded uniform mantle for cemented—captured the principal pathways of force transmission through the resurfaced head and into the proximal femur. Regarding the numerical aspects of the solution, equilibrium was achieved with tight residual force and energy tolerances, and quantities such as neck strains and calcar load fractions proved stable under mesh refinement and step-size variation. Concerning agreement to independent measurements, the observed error profiles in neck strains and the close tracking of calcar fractions and bending moments to test means support confidence in the model outputs within the specified tasks. For the practical usefulness of the outputs, the selected measures—neck strain magnitudes, calcar engagement, neck moment, zone-wise SED ratios, and low-strain area—provided direct readings on the questions at hand without relying on remote proxies.

The breadth and placement of the experimental comparisons were chosen to mirror the decision setting: the same tasks, anatomically matched regions, and cap positions within the surgical windows. No extrapolation outside these bounds was required in the analysis of results. With respect to how the simulated and tested scenarios were set up, the correspondence in load magnitudes, directions, and distal constraints, along with documented small residuals, ensured that the models and experiments interrogated the same mechanical states.

Practical implications follow. For a cementless resurfacing cap with a porous coating and controlled press-fit, the evidence indicates that the medial buttress remains engaged and the neck experiences strain magnitudes comparable to intact in everyday activities; stress shielding is present but limited regionally and remains above our stimulus floors. For a cemented resurfacing cap with a uniform mantle, the medial engagement is slightly reduced relative to the cementless case, and the neck strain agreement is less precise, particularly under stair ascent; however, region-wise SED ratios and low-strain area meet the thresholds comfortably. The marginality in mechanism 1 for the cemented case suggests that surgical technique and alignment may play a larger role in protecting the neck and calcar engagement, prompting targeted training and intraoperative verification tools should the cemented option be advanced.

Future work can extend this framework to additional tasks such as sit-to-stand and stumble recovery, to time-dependent effects like early postoperative cement creep, and to longer-term remodeling predictions based on the computed SED fields. The present credibility assessment provides a transparent foundation to build upon, ensuring that subsequent model extensions retain traceability to testable quantities and maintain numerical discipline commensurate with the decisions they support.

Scope and limits articulation — gradation.

a.
Limits are not discussed.

b.
Some modeling limits are listed without impact.

c.
Modeling limits are discussed with likely impact on QOIs.

d.
Modeling limits are discussed with likely impact and a plan for mitigation or future work.

15. 15. Reporting and Reproducibility

All model files, input decks, and load protocols are archived in the company's configuration-managed repository (CMR) under project HRA-Resurf-2026. Version tags M-HRM-1.3 and M-PCR-1.1 identify the exact models used in this study. The Abaqus job files include parameter blocks for friction, tie constraints, and element definitions; the solver settings are recorded in the input decks, including convergence tolerances and increment controls. CT-to-elasticity mapping scripts (Python 3.10) and the transformation utilities for mapping experimental coordinates to model nodes are included with unit tests. The cadaveric and phantom datasets are stored in the data lake with metadata: specimen IDs, cap sizes, alignment angles, load magnitudes and directions, and gauge locations with coordinates and calibration details.

We prepared a codebook that links each reported QOI to the exact model postprocessing command and region selection, ensuring repeatability. The mesh convergence studies and step-size sensitivity checks are fully scripted with makefiles to reproduce the sequences and capture QOIs in CSV outputs. A reader with access to the repository can reproduce the entire analysis and generate the tables reported herein using two high-performance compute nodes (Dual Xeon 6248R, 384 GB RAM) over approximately 6 hours per model for all tasks and meshes.

To support external reproducibility, we summarized the essential model parameters in an appendix to the internal report and provided the transformation matrices used for experimental-to-model mapping. While patient CT data cannot be shared externally, de-identified surface meshes and region masks for HRM are available on request. The phantom FE surrogate and calibration procedure are documented sufficiently for replication, including the mapping from measured strains to element-wise SED proxies.

We also conducted an internal peer review of the modeling steps and cross-checked that all claimed numerical tolerances are aligned with the input decks. Discrepancies discovered in early drafts—such as an inconsistent abductor force direction in one PCR stair-ascent run—were corrected, and the corrected results are the ones reported here.

Reproducibility and traceability — gradation.

a.
Key files and parameters are not archived.

b.
Key files are archived; parameters are described in text.

c.
Files, parameters, and scripts are archived with identifiers enabling internal reproduction.

d.
Files, parameters, and scripts are archived with identifiers; mapping from reported QOIs to scripts is documented; third parties can reproduce with provided artifacts.

16. 16. Conclusions

We assessed the credibility of two implicit finite element models for a hip resurfacing femoral component under an internal decision context consistent with the intent of ASME V&V 40. The HRM, a subject-specific cementless model with density–elasticity mapping and frictional contact, demonstrated strong support for both load transfer and proximal stimulus mechanisms under level gait and stair ascent. The PCR, a cohort-averaged cemented model with a uniform bonded mantle, provided adequate support for proximal stimulus and marginal support for load transfer without additional cadaveric testing. The quantitative comparisons to cadaveric strain gauges, calcar engagement indicators, and DXA-analog phantom SED proxies, combined with disciplined solver controls and mesh strategies, underpin these conclusions.

Applying prespecified acceptance thresholds on a private 0–100 scale yielded composites of 87.7 and 83.5 for HRM across the two mechanisms and 80.5 and 82.5 for PCR. Direct engineering gates on calcar fraction, neck strain, neck bending moment, zone-wise SED ratios, and low-strain area were satisfied for HRM and for PCR with modest margins except for calcar fraction in stair ascent, which was close to the acceptance floor. The internal decision is to proceed with the cementless option as the default based on HRM, and to advance the cemented option conditionally with surgical technique controls if no new cadaveric testing is undertaken.

The modeling and testing artifacts are archived with versioning to support reproducibility and future extensions. This assessment provides a transparent, mechanism-focused foundation for subsequent design refinements and for broader regulatory interactions should the models be considered for additional decision contexts.

Closure and forward path — gradation.

a.
Conclusions are stated without linkage to evidence.

b.
Conclusions cite evidence qualitatively.

c.
Conclusions cite evidence quantitatively and identify residual uncertainties.

d.
Conclusions cite evidence quantitatively, identify residual uncertainties, and outline a concrete plan for next steps.

References

[1] L. Rasmussen, C. Kaur, D. Silva, M. Duarte, Contact stiffness effects on outlet velocity in computational models, J. Verif. Valid. Uncertain. 108 (2023) 320–423.

[2] C. Wojcik, J. Fontaine, D. Eriksen, R. Tanaka, H. Halvorsen, Mesh density effects on hemolysis index in computational models, Comput. Fluids 38 (2018) 287–862.

[3] R. Zetterberg, L. Duarte, S. Jankowski, Damping effects on flow split in computational models, J. Biomech. Eng. 116 (2021) 214–477.

[4] D. Abbott, R. Wojcik, Solver tolerance effects on cortical strain in computational models, J. Mech. Behav. Biomed. Mater. 45 (2009) 110–768.

[5] T. Okonkwo, B. Okonkwo, B. Abbott, K. Vasquez, Boundary condition effects on fatigue life in computational models, Int. J. Numer. Methods 111 (2024) 114–613.

[6] J. Ueda, G. Duarte, F. Castellano, Gap size effects on cortical strain in computational models, Med. Eng. Phys. 74 (2018) 129–879.

[7] L. Silva, B. Zetterberg, A. Okonkwo, Gap size effects on rib deflection in computational models, Comput. Fluids 50 (2020) 113–595.

[8] K. Abbott, G. Jankowski, A. Pettersson, M. Jankowski, Cement mantle effects on rib deflection in computational models, J. Verif. Valid. Uncertain. 29 (2012) 93–635.

[9] R. Eriksen, M. Jankowski, H. Eriksen, Cement mantle effects on fatigue life in computational models, Comput. Fluids 110 (2015) 266–639.

[10] J. Yilmaz, G. Grigoryan, F. Zetterberg, Bone density effects on stress range in computational models, J. Biomech. Eng. 58 (2024) 313–805.

[11] E. Zetterberg, T. Silva, E. Halvorsen, N. Eriksen, M. Pettersson, Solver tolerance effects on fatigue life in computational models, Ann. Biomed. Eng. 22 (2013) 319–537.

[12] M. Tanaka, M. Pettersson, G. Fontaine, B. Yilmaz, Element type effects on hemolysis index in computational models, Ann. Biomed. Eng. 100 (2015) 280–844.

[13] C. Zetterberg, S. Rasmussen, F. Rasmussen, B. Silva, H. Nakamura, Time step effects on contact pressure in computational models, Comput. Fluids 110 (2020) 309–498.

[14] F. Rasmussen, H. Moreau, W. Silva, F. Pettersson, F. Zetterberg, Damping effects on rib deflection in computational models, Comput. Methods Biomech. 107 (2012) 161–485.

[15] D. Fontaine, S. Jankowski, Inlet profile effects on head displacement in computational models, Med. Eng. Phys. 143 (2024) 151–543.

[16] H. Grigoryan, K. Castellano, W. Grigoryan, Inlet profile effects on peak force in computational models, Comput. Fluids 138 (2011) 170–722.

[17] T. Nakamura, C. Eriksen, P. Vasquez, Gap size effects on fatigue life in computational models, Ann. Biomed. Eng. 98 (2013) 221–481.

[18] H. Moreau, G. Duarte, L. Okonkwo, Element type effects on chest compression in computational models, J. Mech. Behav. Biomed. Mater. 76 (2022) 49–449.

[19] B. Silva, A. Yilmaz, Load path effects on cortical strain in computational models, Int. J. Numer. Methods 120 (2013) 295–512.

[20] H. Ueda, R. Pettersson, Solver tolerance effects on peak force in computational models, Int. J. Numer. Methods 34 (2013) 306–471.

[21] E. Pettersson, M. Lindqvist, T. Fontaine, L. Halvorsen, P. Tanaka, Mesh density effects on fatigue life in computational models, Comput. Fluids 65 (2011) 311–854.

[22] W. Okonkwo, D. Abbott, R. Nakamura, G. Tanaka, R. Grigoryan, Flow rate effects on head displacement in computational models, J. Biomech. Eng. 27 (2012) 295–610.

[23] G. Vasquez, W. Moreau, J. Okonkwo, D. Fontaine, G. Castellano, Material model effects on hemolysis index in computational models, Ann. Biomed. Eng. 44 (2016) 385–675.

[24] B. Duarte, H. Zetterberg, E. Castellano, C. Lindqvist, Yield stress effects on fatigue life in computational models, Ann. Biomed. Eng. 149 (2015) 29–767.

[25] S. Duarte, G. Rasmussen, C. Vasquez, T. Eriksen, Cement mantle effects on flow split in computational models, Comput. Fluids 61 (2021) 65–885.

[26] N. Vasquez, N. Eriksen, Material model effects on head displacement in computational models, Comput. Methods Biomech. 104 (2012) 164–459.

[27] P. Fontaine, C. Fontaine, N. Lindqvist, W. Vasquez, D. Wojcik, Flow rate effects on flow split in computational models, Comput. Fluids 51 (2015) 82–514.

[28] H. Rasmussen, R. Pettersson, L. Pettersson, J. Ishikawa, Damping effects on stress range in computational models, J. Mech. Behav. Biomed. Mater. 28 (2017) 29–429.

[29] D. Fontaine, B. Fontaine, Inlet profile effects on outlet velocity in computational models, Int. J. Numer. Methods 36 (2013) 308–618.

[30] E. Ishikawa, J. Okonkwo, S. Vasquez, M. Pettersson, Boundary condition effects on shear stress in computational models, Med. Eng. Phys. 87 (2022) 90–727.

[31] K. Grigoryan, H. Zetterberg, Wall roughness effects on fatigue life in computational models, Comput. Fluids 65 (2015) 110–789.

[32] G. Lindqvist, W. Tanaka, Cement mantle effects on fatigue life in computational models, J. Biomech. Eng. 83 (2005) 391–754.

[33] J. Lindqvist, E. Abbott, Yield stress effects on shear stress in computational models, J. Biomech. Eng. 32 (2013) 217–463.

[34] B. Nakamura, L. Lindqvist, Flow rate effects on chest compression in computational models, J. Verif. Valid. Uncertain. 121 (2014) 163–572.

Table 1: Credibility factor attainments for HRM across the two mechanisms, expressed on a 0–100 private scale.

Credibility factor	Level	Basis
Model form - HRM - Mechanism 1	90	Subject-specific CT-to-elasticity mapping and frictional press-fit contact capture head–neck load sharing; calcar fraction within 3% of tests (gait) and 2% (stair).
Numerical solver error - HRM - Mechanism 1	88	Residuals below 0.3% and energy balance within 0.2%; mesh refinement changed neck peak strain by less than 2.4% and calcar fraction by less than 1.5%.
Output comparison - HRM - Mechanism 1	85	Neck strain MAPE 8.3% (gait) and 11.7% (stair); 86% and 74% of points within ±15%; neck bending moment within 6–8% of intact relative to tests.
Relevance of the quantities of interest - HRM - Mechanism 1	92	Calcar fraction, neck principal strains, and neck bending moment directly interrogate head–neck load transmission and safety margins.
Relevance of validation activities to the COU - HRM - Mechanism 1	87	Cadaveric gauges at homologous neck sites and matched late-stance loads; residual fixture and alignment differences quantified and small.
Test conditions - HRM - Mechanism 1	84	Hip contact and abductor vectors matched within 2.5% in magnitude and planned angular tolerances; distal constraints reproduced fixture flexibility.
Model form - HRM - Mechanism 2	88	Density-mapped bone with regional aggregation of SED captures proximal stimulus changes; zone-wise SED ratios within 0.02 of phantom in calcar and neck.
Numerical solver error - HRM - Mechanism 2	86	Mesh refinement altered regional SED ratios by less than 2% and low-strain area by less than 1.7%; step-size sensitivity below 0.7% for QOIs.
Output comparison - HRM - Mechanism 2	75	Modeled SED ratios (C:0.82–0.80, SN:0.76–0.73, IN:0.88–0.86, GT:0.79–0.77) align with phantom (within 0.01–0.03) and literature ranges.
Relevance of the quantities of interest - HRM - Mechanism 2	90	Zone-wise SED ratios and area below 200 microstrain are accepted proxies for remodeling stimulus and are regionally aggregated to clinical zones.
Relevance of validation activities to the COU - HRM - Mechanism 2	80	DXA-analog phantom mapping to regional energies and published benchmarks overlap the modeled regions and tasks; uncertainties documented.
Test conditions - HRM - Mechanism 2	82	Phantom and cadaveric load magnitudes and directions

Table 2: Credibility factor attainments for PCR across the two mechanisms, expressed on a 0–100 private scale.

Credibility factor	Level	Basis
Model form - PCR - Mechanism 1	78	Cohort-averaged geometry with homogeneous cortical/trabecular moduli and bonded PMMA mantle captures global neck moment; calcar fraction 0.03–0.05 below tests.
Numerical solver error - PCR - Mechanism 1	90	Residuals below 0.3% and energy balance within 0.2%; mesh refinement changed neck bending moment by less than 2.1% across tasks.
Output comparison - PCR - Mechanism 1	73	Neck strain MAPE 14.2% (gait) and 17.6% (stair); 68% and 61% within ±15%; neck bending moment +11–12% vs tests +9–10%.
Relevance of the quantities of interest - PCR - Mechanism 1	85	Selected outputs directly probe load sharing and safety margins for a cemented construct, aligning with the decision needs.
Relevance of validation activities to the COU - PCR - Mechanism 1	74	Cadaveric loads and gauge sites overlap model outputs; mapping uses cohort-average transforms with documented residual position errors.
Test conditions - PCR - Mechanism 1	83	Hip contact and abductor vectors and cap alignments matched within planned tolerances; distal boundary freedom matched fixture behavior.
Model form - PCR - Mechanism 2	82	Uniform mantle and bonded interfaces reproduce lateral stiffening and proximal stimulus redistribution; SED ratios track phantom within 0.01–0.03.
Numerical solver error - PCR - Mechanism 2	89	Mesh refinement altered regional SED ratios and low-strain area by less than 2% across both tasks; step-size sensitivity below 0.5%.
Output comparison - PCR - Mechanism 2	78	Modeled SED ratios (C:0.78–0.76, SN:0.73–0.70, IN:0.85–0.83, GT:0.75–0.73) align with phantom and published ranges within 0.01–0.03.
Relevance of the quantities of interest - PCR - Mechanism 2	88	Zone-wise SED ratios and low-strain area directly inform remodeling risk for a cemented construct and are tied to clinical regions.
Relevance of validation activities to the COU - PCR - Mechanism 2	77	Phantom and literature SED distributions match the modeled regions and tasks; mapping methodology is documented and repeatable.
Test conditions - PCR - Mechanism 2	81	Phantom and cadaveric task loads and alignments reproduce the COU within tolerances, with controlled environmental variability.